



Original citation:

Langley, K., Fleet, D. J. and Atherton, T. J. (1991) An instantaneous frequency-based computation of transparent motion. University of Warwick. Department of Computer Science. (Department of Computer Science Research Report). (Unpublished)

Permanent WRAP url:

<http://wrap.warwick.ac.uk/60887>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

A note on versions:

The version presented in WRAP is the published version or, version of record, and may be cited as it appears here. For more information, please contact the WRAP Team at: publications@warwick.ac.uk



<http://wrap.warwick.ac.uk/>

Research Report 200

An Instantaneous Frequency-Based Computation of Transparent Motion

K. Langley, D.J. Fleet, T.J. Atherton

RR200

We extend the principle of phase-based techniques for measuring optical flow and binocular disparity to multiple motion estimation. Using quadrature filter pairs, we estimate phase gradients (instantaneous frequency) from independent bandpass channels, which we apply towards the problem of multiple optic flow analysis. Our approach is similar to that of Shizawa and Mase [23], in which n th-order differential operators are required to compute n simultaneous velocity estimates. The approach presented here is simpler, however, because we require only a set of band-pass filters, the out-puts of which need only be differentiated once. Thus, the present technique requires only first-order filters. Additionally, this theory is also applicable to transparency in stereopsis.

An Instantaneous Frequency-Based Computation of Transparent Motion

K. Langley

Department of Experimental Psychology,
University College London,
London, U.K. WC1E 6BT

D. J. Fleet

Department of Computing Science,
Queen's University,
Kingston, Ontario, Canada K7L 3N6

T. J. Atherton

Department of Computer Science,
University of Warwick,
Coventry, U.K. CV4 7AL

Abstract

We extend the principle of phase-based techniques for measuring optical flow and binocular disparity to multiple motion estimation. Using quadrature filter pairs, we estimate phase gradients (instantaneous frequency) from independent bandpass channels, which we apply towards the problem of multiple optic flow analysis. Our approach is similar to that of Shizawa and Mase [23], in which n th-order differential operators are required to compute n simultaneous velocity estimates. The approach presented here is simpler, however, because we require only a set of band-pass filters, the outputs of which need only be differentiated once. Thus, the present technique requires only first-order filters. Additionally, this theory is also applicable to transparency in stereopsis.

1 Introduction

Within motion analysis, there are two distinct points of view. There are techniques that detect and track edges over time (e.g [5]), and those that compute explicit measurements from the image intensity pattern. Edge-based approaches must typically deal with aperture problem. For intensity-based approaches this only becomes a problem where the image intensity function is comprised mainly of a single one-dimensional structure. In general, intensity-based approaches make better use of the entire input signal.

Intensity-based approaches can themselves be subdivided into three different groups. Spatio-temporal energy models [11, 3, 1] compute image velocity from the relative amplitudes of the outputs of different band-pass filters. This approach, however, does not perform well when all of the power of the signal lies in the passband of a single filter. It is well-known in human vision [2] that the human visual system performs extremely well under these simple conditions. There are also differential techniques that measure velocity from spatiotemporal derivatives of intensity of band-pass filter outputs (e.g. [24]). However, these techniques can be sensitive to noise, geometric deformations between frames, and photometric variations as they assume conservation of image intensity, or its filtered representation [7].

The approach taken here is phase-based [14, 25, 17, 15]. Measurements are based on a representation of the signal structure provided by a family of quadrature-pair bandpass filters. The convolution of an image with a linear bandpass operator is given by

$$R(\mathbf{x}, t) = I(\mathbf{x}, t) * K(\mathbf{x}, t) = \rho(\mathbf{x}, t) \exp[i\phi(\mathbf{x}, t)] \quad (1)$$

where $K(\mathbf{x}, t)$ is a complex-valued bandpass kernel, and $\rho(\mathbf{x}, t)$ and $\phi(\mathbf{x}, t)$ are the amplitude and phase components of the bandpass response. We are assuming that the filters are effectively quadrature pairs and can be expressed as the product of a lowpass envelope and a complex exponential, for example, Gabor filters.

Interestingly, differentiation of the above equation provides two independent equations that can in principle be applied to solve the aperture problem from the output of a single filter [18]. It should be noted that similar envelope/phase properties are also obtained from differential operators, but the traditional representation of these filter kernels does not pay attention to the bandpass signal. We first consider the motion constraint equation [12] applied to the bandpass signal, assuming $\frac{dR(\mathbf{x}, t)}{dt} = 0$:

$$\exp[i\phi(\mathbf{x}, t)] [(\rho_x + i\phi_x\rho)v_1 + (\rho_y + i\phi_y\rho)v_2 + (\rho_t + i\phi_t\rho)] = 0 \quad (2)$$

where subscripts refer to the direction of partial differentiation with respect to the coordinate frame. This gives two equations:

$$\mathcal{Im} \frac{d \ln R(\mathbf{x}, t)}{dt} = \phi_x v_1 + \phi_y v_2 + \phi_t = 0 \quad (3)$$

$$\mathcal{Re} \frac{d \ln R(\mathbf{x}, t)}{dt} = \frac{\rho_x}{\rho} v_1 + \frac{\rho_y}{\rho} v_2 + \frac{\rho_t}{\rho} = 0 \quad (4)$$

resulting in the following velocity measurement for single motion flow:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \phi_x & \phi_y \\ \rho_x & \rho_y \end{bmatrix}^{-1} \begin{bmatrix} \phi_t \\ \rho_t \end{bmatrix} \quad (5)$$

that is, velocity can be measured using a general representation of the bandpass signal. In [16] temporal energy derivatives and phase derivatives were deliberately kept independent because the independence of amplitude derivatives from phase derivatives provides an account for transparency and coherence in human vision. However, in practice we tend to rely on the phase part of the signal because of stability [7] considerations. Hence two independent bandpass filters are required to compute an unambiguous velocity vector in the case of single image flow.

The recognition that the human visual system is capable of separately analysing several independent motions at the same point in the image domain, has prompted some authors[21, 4] to investigate the computational rationale behind multiple optical flow analysis. It is hoped that algorithms suitable for measuring multiple image velocities in a single image neighbourhood will be helpful in a wide variety of circumstances, including the superposition of signals, nonlinear transparent phenomena, and several forms of occlusion. Towards this end, several methods have been proposed based on the superposition of two or more translating signals (e.g. [21, 23, 4, 13]). This paper discusses a variation on this theme, using the previous work of Shizawa and Mase[21, 23] as a starting point. It is shown that multiple motions may be computed without the need to estimate second or higher-order derivatives. The problem of computing multiple motion is posed instead in terms of constraints on local measures of instantaneous frequency from different band-pass filter outputs. With a preliminary implementation we find that this method produces reliable estimates of two simultaneous motions of superimposed signals. It also appears to be robust with respect to multiplicative combinations of translating signals, and differences in signal power with respect to transparent surfaces.

The occurrence of more than one legitimate image velocity in a single image neighbourhood may be caused by one of several common phenomena:

- specularities or mirror-like surface reflections like those off a polished floor;
- shadows under diffuse lighting conditions that are seen to move across a stationary surface;
- occlusion such as a single occluding boundary or the fragmented occlusion caused by natural vegetation or certain fences;
- translucency, in which light reflected from one surface is passed through another to the camera, such as stained (or dirty) glass;
- and atmospheric phenomena such as smoke, rain or snow.

In just the past few years several methods that address the problem of multiple image velocities have emerged. For example, some methods compute velocity histograms in relatively local regions of the image [10, 13]. Similarly, Fleet and Jepson[8] showed that local phase information can be used to compute multiple estimates of the normal component of 2-d velocity. However, these techniques do not address the segmentation of the different local measurements to compute separate 2-d velocity estimates. Langley and Fleet[16] have argued that the independence of phase and energy velocity does provide a basis to explain transparent motion to simple signals in human vision. They also noted that the group (energy) velocity of the image signal is not constrained to pass through the origin of the frequency domain, which is one of the properties of multiplicative motion transparency. Bergen et al[4], have derived an iterative method that initially locks onto the one of the motions, allowing it to be cancelled by substrating a deformed version of one frame from another. The same operations can then be applied to the resulting sequence to detect other motions that might exist. However, only Shizawa and Mase[21, 23] have attempted to obtain explicit constraint equations for the analysis of multiple flows from image sequences. Their approach requires that second or higher order derivatives be extracted in space and time, and averaged throughout local apertures in order to estimate the parameters of motion. In order to compute n image velocities simultaneously requires the application of n^{th} -order differential operators.

By contrast, the approach presented here recasts the multiple-flow motion constraint equation in terms of a constraint on instantaneous frequencies of the signals. We view this as an extension to phase-based methods for computing image velocity and binocular disparity, in which image velocity and binocular disparity are measured from the output of band-pass filters [14, 25, 17, 15]. Here we are specifically concerned with the use of the phase gradient, which provides a measure of the instantaneous frequency of the filter response as a function of space and time. Instantaneous frequency may be computed from the filter outputs directly, without explicitly representing the phase signal. In addition, we note that the specific form of band-pass filters is not crucial to the approach. Moreover, there exist recent results concerning the general stability of phase information, as well as its potential instabilities, that we may exploit to use instantaneous frequency in a reliable manner [14, 7].

With the use of instantaneous frequency we avoid the need to estimate higher-order derivatives which are typically noise sensitive. Furthermore, there is a straightforward simplification of this technique to the two-frame case of stereopsis. It is well known[19] that several planes can be observed by human observers using stereo pairs of random dots. Since with two sensors, it is only possible to apply a discrete approximation to a first order

derivative, we should investigate the possibility of computing multiple motions without second derivatives. We find that the technique presented here can be applied to measure surfaces at multiple disparities under certain conditions, as well as multiple image velocities.

2 Background Theory

With respect to motion transparency, much of the groundwork has already been covered by Shizawa and Mase[21, 23, 22] in terms of understanding the necessary constraints that are required to derive several motions from an image intensity function. They have also shown the relationship between their work, and that of Bergen et al[4].

The results of Shizawa and Mase[21, 23, 22] are based on the superposition of two translating signals. In this case, not only does the motion constraint equation apply to each of the component signals, but there is a combined constraint that applies to their superposition. For example, let $f(\mathbf{x}, t)$ be the sum of two translating signals, $f_1(\mathbf{x}, t)$ and $f_2(\mathbf{x}, t)$, with velocities $\mathbf{v}_1 = (u_1, v_1)$ and $\mathbf{v}_2 = (u_2, v_2)$. Individually, the signals satisfy the motion constraint equations:

$$(\mathbf{v}_j, 1) \cdot \nabla f_j(\mathbf{x}, t) = 0, \quad j = 1, 2, \quad (6)$$

where $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial t}]$. Their superposition $f(\mathbf{x}, t)$ then satisfies:¹

$$((\mathbf{v}_1, 1) \cdot \nabla) ((\mathbf{v}_2, 1) \cdot \nabla) f(\mathbf{x}, t) = 0, \quad (7)$$

where $(\mathbf{v}, 1) \cdot \nabla = [u \frac{\partial}{\partial x}, v \frac{\partial}{\partial y}, \frac{\partial}{\partial t}]$. When (7) is expanded, the individual terms are found to be :

$$u_1 u_2 f_{xx} + v_1 v_2 f_{yy} + (u_1 v_2 + u_2 v_1) f_{xy} + (u_1 + u_2) f_{xt} + (v_1 + v_2) f_{yt} + f_{tt} = 0. \quad (8)$$

Using differential measurements of $f(\mathbf{x}, t)$ at four or five points Shizawa and Mase [22] describe how to compute the individual 2-d velocities. The fitting of three velocity planes through the origin of the frequency domain is a direct extension of this formalism to include higher-order differential operators.

3 Constraints on Instantaneous Frequency

It is well-known that the translation of a 2-d pattern has all its power concentrated on a plane in the frequency domain [6]; that is, the Fourier transform of (6) satisfies:

$$\hat{f}_j(\mathbf{k}, \omega) = \hat{h}(\mathbf{k}) \delta(\mathbf{v}_j \cdot \mathbf{k} + \omega), \quad (9)$$

¹This derivation assumes more than the conservation of f_1 and f_2 , as would be required by (6) alone. In (7), because of the cascaded differentiation, it is important that the two velocities \mathbf{v}_1 and \mathbf{v}_2 be constant as a functions of space and time.

where \mathbf{k} and ω are spatial and temporal frequency variables, $\delta(\cdot)$ is a Dirac delta function, and $\hat{h}(\mathbf{k})$ represents the 2-d Fourier transform of the 2-d pattern that is translating. The velocity constraint in (9) is:

$$\mathbf{v}_j \cdot \mathbf{k} + \omega = 0 , \quad (10)$$

which also follows from the Fourier transform of (6).

In these terms, finding a solution to (8) for the two velocities amounts to simultaneously fitting two planes to the power of $f(\mathbf{x}, t)$ in the frequency domain. Towards this end, note that the Fourier transform of (8) is given by:

$$\begin{aligned} iu_1u_2k_1^2\hat{f}(\mathbf{k}, \omega) + iv_1v_2k_2^2\hat{f}(\mathbf{k}, \omega) + i(u_1v_2 + u_2v_1)k_1k_2\hat{f}(\mathbf{k}, \omega) + \\ i(u_1 + u_2)k_1\omega\hat{f}(\mathbf{k}, \omega) + i(v_1 + v_2)k_2\omega\hat{f}(\mathbf{k}, \omega) + i\omega^2\hat{f}(\mathbf{k}, \omega) = 0 . \end{aligned} \quad (11)$$

If we factor out $i\hat{f}(\mathbf{k}, \omega)$ from (12), we are left with the constraint:

$$u_1u_2k_1^2 + v_1v_2k_2^2 + (u_1v_2 + u_2v_1)k_1k_2 + (u_1 + u_2)k_1\omega + (v_1 + v_2)k_2\omega + \omega^2 = 0 . \quad (12)$$

In effect, (12) constrains the locations of nonzero power in the frequency domain (\mathbf{k}, ω) to lie on one of two planes. Appendix A provides another perspective on the constraints in (12), in order to show the similarity of this constraint to that of Shizawa and Mase.

Our approach to solving for the two velocities involves finding a solution to the coefficients in (12) that contain the components of \mathbf{v}_1 and \mathbf{v}_2 . We then compute the individual velocities from these terms. For convenience, we rewrite (12) directly in vector form as:

$$\mathbf{a}^T \mathbf{m} = 0 , \quad (13)$$

where $\mathbf{m} = (k_1^2, k_2^2, k_1k_2, k_1\omega, k_2\omega, \omega^2)^T$, and $\mathbf{a} = (u_1u_2, v_1v_2, u_1v_2 + u_2v_1, u_1 + u_2, v_1 + v_2, 1)^T$. At least 5 measurements of instantaneous frequency (\mathbf{k}_j, ω_j) (at which there is significant power) are required to solve for the five unknown elements of \mathbf{a} . Given six or more measurements of instantaneous frequency we have an overconstrained system, and can solve for the elements of \mathbf{a} more robustly. We do this by minimizing the squared error between the model and the instantaneous frequencies; that is, we minimize

$$\sum_j (\mathbf{a}^T \mathbf{m}_j)^2 \quad (14)$$

with respect to \mathbf{a} . Differentiating (14) with respect to \mathbf{a} , and setting the result to zero produces the linear normal equations:

$$M \mathbf{a} = \mathbf{0} , \quad \text{where } M \equiv \sum_j \mathbf{m}_j \mathbf{m}_j^T . \quad (15)$$

Equation (15) constrains \mathbf{a} to lie in the null space of M , the matrix of outer products. Ideally, in the case of two motions, M has a rank of 5, with a 5-dimensional column space and a 1-dimensional null space. The null space is spanned by the eigenvector corresponding to the zero eigenvalue. Therefore, to solve for the elements of \mathbf{a} , we compute the smallest eigenvalue of M (which should be zero for two motions), and its corresponding eigenvector. We then scale the eigenvector so that its last element is unity, which produces our least squares estimate of \mathbf{a} . An alternative approach based upon the method of the pseudo-inverse is shown in Appendix B.

From the elements of \mathbf{a} , as described in [22], the individual velocities are found as follows: For convenience, let the computed elements of \mathbf{a} be denoted a_j , $j = 1, \dots, 5$. Then, the two x -components of velocity are given by solutions to:

$$u_j = \frac{1}{2}a_4 \pm \sqrt{\frac{1}{4}a_4^2 - a_1}, \quad (16)$$

while the two y -components of velocity are given by:

$$v_j = \frac{1}{2}a_5 \pm \sqrt{\frac{1}{4}a_5^2 - a_2}. \quad (17)$$

The correct combinations of these roots to one another to obtain estimates of \mathbf{v}_1 and \mathbf{v}_2 are then determined by a_3 . In particular, note that there are only two ways to combine the different estimates of velocity in the x and y directions. Only one of these two will equal the third component of \mathbf{a} .

4 Computing Instantaneous Frequency

In order to compute various measurements of instantaneous frequency, we assume that there exist a family of band-pass filters that are initially applied to the image sequence, such as those used by Heeger, or Fleet and Jepson [11, 6]. If the tuning of the filters is sufficiently different, we can assume that the frequency measurements represent independent degrees of freedom of instantaneous frequency measurements.

Let $K(\mathbf{x}, t; \mathbf{k}_0, \omega_0)$ be a complex-valued band-pass kernel with peak tuning frequency (\mathbf{k}_0, ω_0) , and let $R(\mathbf{x}, t; \mathbf{k}_0, \omega_0)$ denote the filter output when applied to the input $f(\mathbf{x}, t)$. That is, R is the convolution of K and f :

$$R(\mathbf{x}, t; \mathbf{k}_0, \omega_0) = K(\mathbf{x}, t; \mathbf{k}_0, \omega_0) * f(\mathbf{x}, t). \quad (18)$$

Because the filter and its output are complex-valued, we may express $R(\mathbf{x}, t; \mathbf{k}_0, \omega_0)$ as:

$$R(\mathbf{x}, t; \mathbf{k}_0, \omega_0) = \rho(\mathbf{x}, t; \mathbf{k}_0, \omega_0) \exp[i \phi(\mathbf{x}, t; \mathbf{k}_0, \omega_0)], \quad (19)$$

where ρ and ϕ are referred to as its amplitude and phase components, which are functions of space-time.

Instantaneous frequency is defined as the spatiotemporal phase gradient [9, 20]. In effect, the instantaneous frequency provides us with a local approximation to the structure of the filter response in terms of an amplitude-modulated, sinusoidal signal. We measure the instantaneous frequency of the filter output R using the identity:

$$\phi_x(\mathbf{x}, t) = \frac{\text{Im}[R^*(\mathbf{x}, t) R_x(\mathbf{x}, t)]}{|R(\mathbf{x}, t)|^2}, \quad (20)$$

where R^* is the complex conjugate of R . It is a simple matter to differentiate the filter output.

But not all phase gradients are useful in constraining the multiple motions that may exist in the image. First, it is important that more weight be given to those frequencies that correspond to greater amounts of local energy, given by the amplitude of the filter output $|R|$. Second, it is important that the measurements of instantaneous frequency be ignored in regions where the phase of the filter output is overly sensitive to small variations in spatial position of the scale of input. These are detected using the theory of phase singularities described by Jepson and Fleet [14, 6]. They occur because of interference between energy maxima in the power spectrum. Finally, we only expect to be able to isolate transparent motion when there is some parameter (scale, orientation or speed) that can be used to distinguish the different motions. We require at least 5 instantaneous frequencies from filters that respond primarily to only one of the two motions.

5 Preliminary Implementation and Results

We have completed a preliminary implementation of this approach that works for 1-d and 2-d signals. In the former case, 1-d signals provide the opportunity to display phase gradients as images, and hence an intuitive grasp of the approach: that is, we are relying on the properties of the bandpass filter to discriminate the phase velocity of transparent motion fields. The details are as follows: In figure 1 a two dimensional pattern consisting of the summation of two one dimensional translating noise sequences in time is shown. The noise sequence has been filtered with a quadrature filter pairs (Gabor functions) tuned to the direction of the motion patterns. The instantaneous frequency is simply the gradient of the phase function.

Results are shown for the translating noise patterns shown in figures 2a and b. The orientation difference between flow fields was 0.49 radians. For the case of additive noise with the two signals of equal contrast, the mean error in velocity estimation from the actual

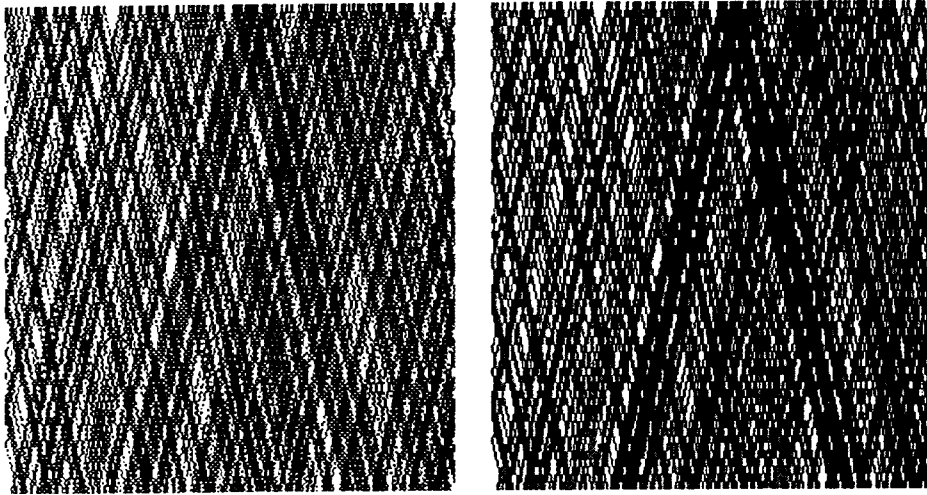


Figure 1: *a) Top Left Additive Transparency for two 1-d noise signals. b) Top Right Multiplicative transparency.*

flow field was found to be -0.00570 radians for leftward motion, and -0.000797 radians for rightward motion. The standard deviation of errors in velocity was measured at 0.02841 and 0.02766 radians respectively. We also show the energy response from the filters used in figure 2a. Notice that the energy term is sensitive to interference from the other flow field, which is particularly true when the magnitude of the energy response is low. This result provides further evidence for the stability of phase information over energy[14] in the context of image analysis. Figure 3 shows phase contours after reducing the contrast of the leftward moving noise pattern of figure 1a to 30% by comparison to the rightward pattern. The mean velocity error for the stronger signal was found to be -0.000697 as opposed to -0.002009 radians, with a standard deviation of 0.01617 in contrast to 0.06381 radians. It is clear, that a reduction in contrast does affect the stability and accuracy of the measurement process. The final example in 1-d is that of multiplicative motion transparency (1b and 2b). In this case we have obtained the poorest results with (this is not surprising since the multiplication of two noise patterns is equivalent to the convolution of the Fourier spectra of the two signals in the frequency domain, and hence highly unrepresentative of the original patterns) a mean error of 0.006944 and 0.000093 radians, and standard deviation of 0.0696 and 0.0891 radians respectively for leftward and rightward motion.

The final example in 7 shows the velocity and error field in the case of 2-d motion, with two added random noise patterns moving independently to each other. Velocity flow fields are presented in diagrammatic form in figures 5 to 7.

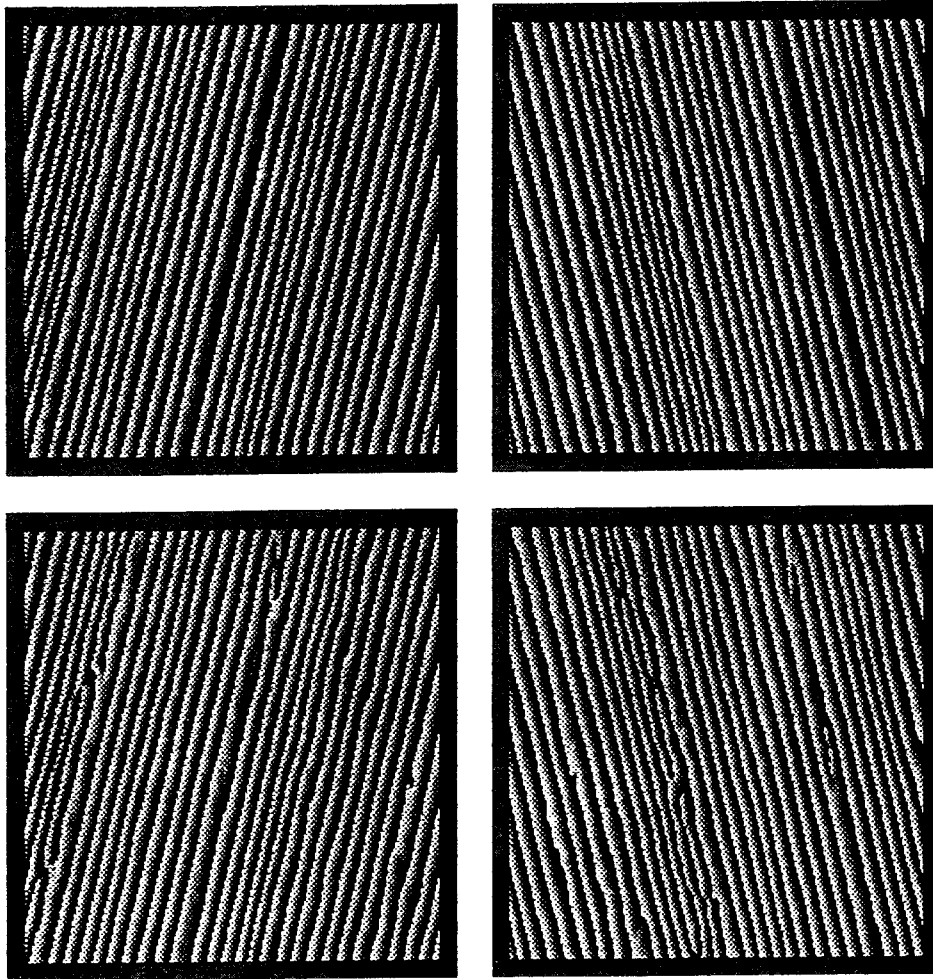


Figure 2: *a) Top Phase contours for two oriented filters for additive transparency. b) Bottom Phase contours for multiplicative transparency taken from figure 1*

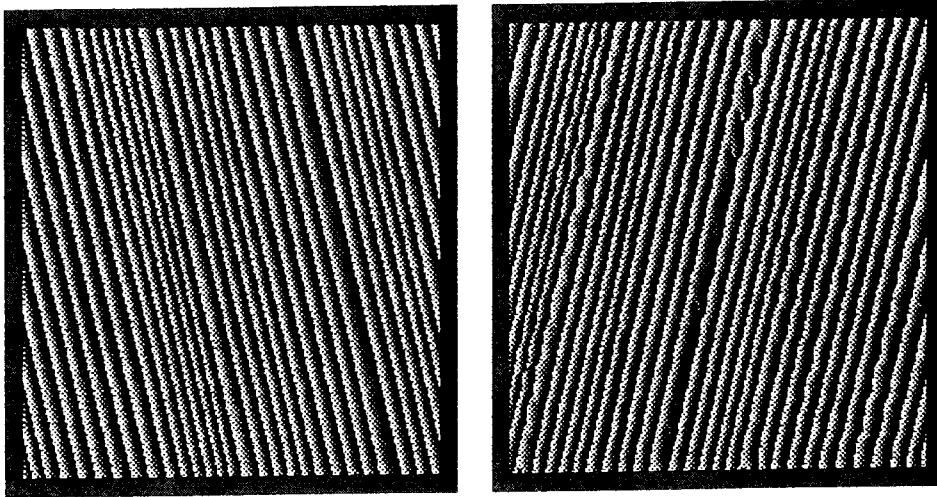


Figure 3: *Phase contours for two oriented filters for additive transparency taken from figure 2a with the contrast of the pattern moving to the left at 30% to that of the pattern moving to the right.*

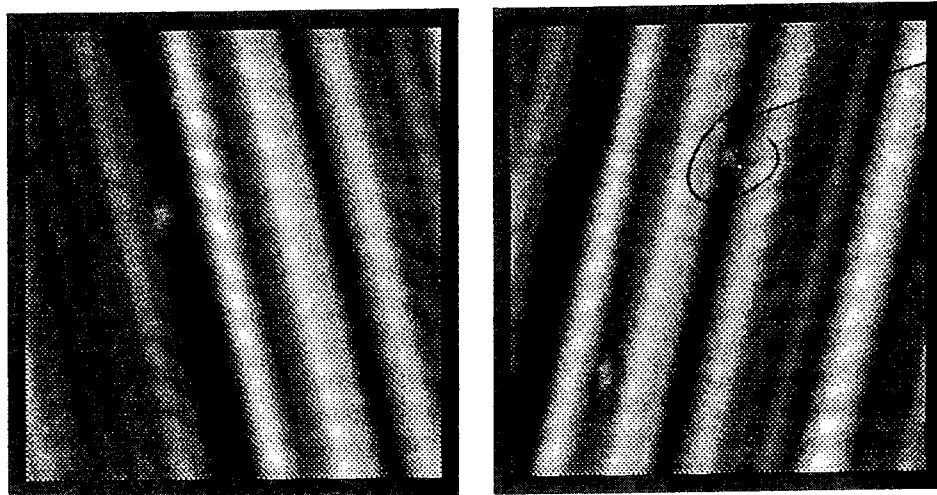


Figure 4: *Energy response from filters producing the phase contours of figure 2a.*

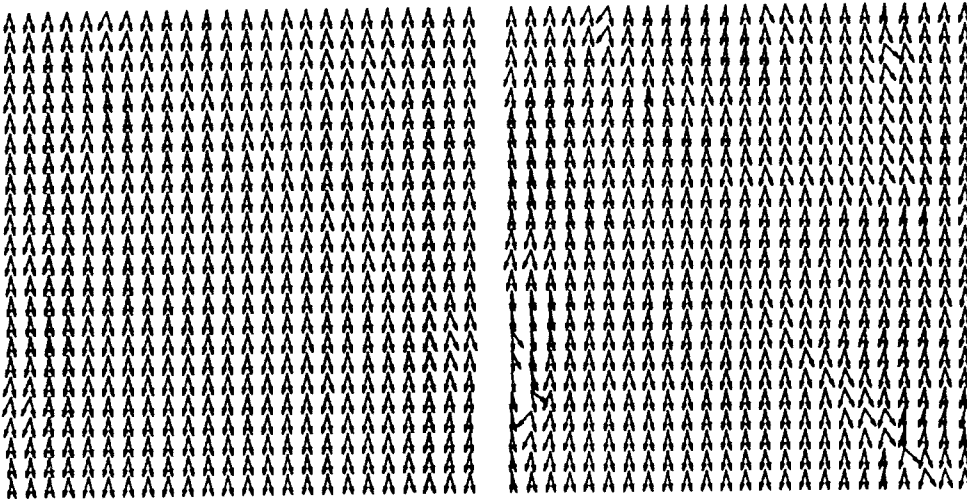


Figure 5: *Needle diagram of velocity flows for figure 2a and 2b. Velocity errors are given in the text.*

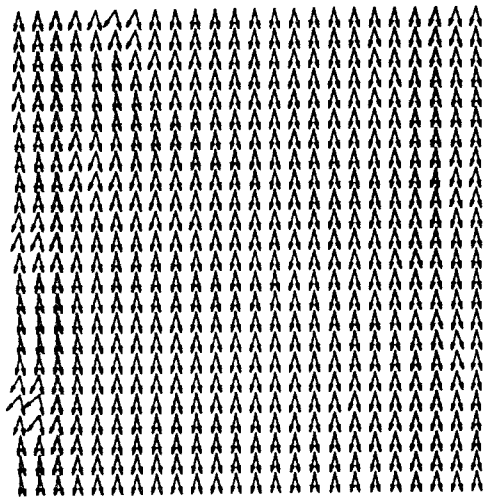


Figure 6: *Needle diagram of velocity flows for figure 2a with the pattern translating to the left superimposed with a contrast ratio of 1:3 to that of the pattern moving to the right.*

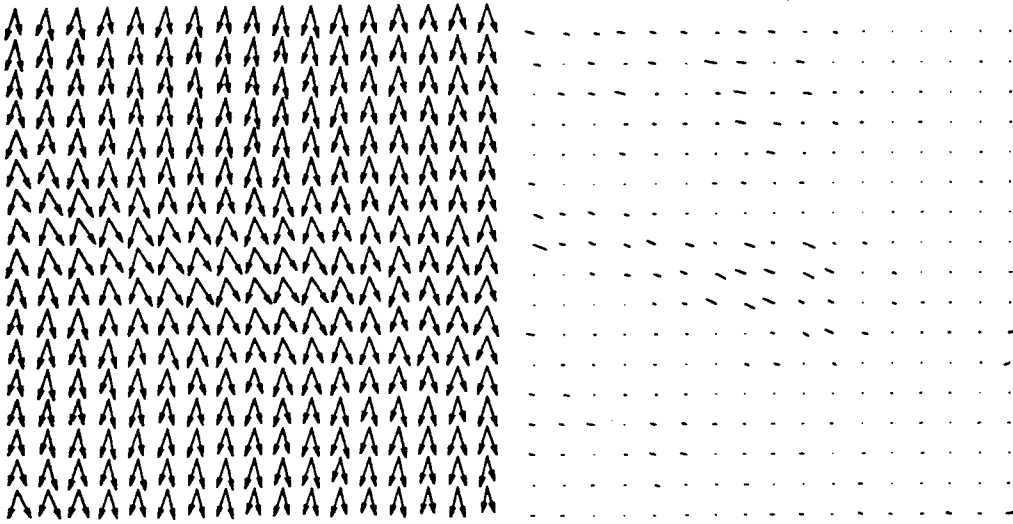


Figure 7: a) Needle diagram for velocity flows for two random noise fields of equal contrast with independent motion. b) Error vectors.

6 Conclusion

This paper outlines a new method for computing multiple optical flows using quadrature-pair filters and their first-order derivatives. The approach is extendable to several independent velocities by increasing the number of filters and modifying our constraint equation. The basic approach is viewed a variation on the theme discussed in detail by Shizawa and Mase. But it offers a substantially different perspective, since it requires only first-order filters, and a mechanism to estimate instantaneous frequency.

There are a number of advantages in the approach that we have chosen. In particular, our processing paradigm allows higher order (deformation, dilation, rotation) properties of the optic flow field to be derived from further differentiation of the bandpass signal representation. Additionally, our preliminary results indicate that our theory can also be applied to the case of stereopsis. This direct extension of our technique is possible because we restrict our processing to first order operators, and hence obtain a discrete approximate to differentiation. Furthermore, it is hoped that the multiple velocity estimates can be made more robustly without requiring second or higher order differentiation of the image. Shizawa and Mase[23] also proposed a theory to determine the multiplicity of image flow. We have not approached this important aspect, since the main point of our paper is to highlight the reduction in differential order that is possible by phase based analysis.

Finally, although mentioned in the introduction, we have not discussed the role of the energy derivative in the context of multiple flows. At present we are unsure about the stability of this parameter with respect to real image data, and this remains a question for future research.

References

- [1] Adelson.E.H and Bergen.J.R. Spatiotemporal energy models for the perception of motion. *J.Opt. Soc. Am. A*, 2,2:284–299, 1985.
- [2] Adelson.E.H. and Movshon.J.A. Phenominal coherence of moving visual patterns. *Nature*, 300:523–525, 1982.
- [3] Atherton.T.J. The detection and measurement of visual motion. *Pattern Recognition Letters*, 5,2:157–163, 1987.
- [4] Bergen.J.R, Burt.P, Hingorani.R, and Peleg.S. Computing two motions from three frames. *Proc. 3rd ICCV, Osaka, Japan*, pages 27–32, 1990.
- [5] Deriche.R and Faugeras.O. Tracking line elements. In Faugeras.O, editor, *Computer Vision-ECCV 90*, pages 341–345, Antibes,France, 1989. Springer-Verlag.
- [6] Fleet.D.J. PhD thesis, University of Toronto, 1990.
- [7] Fleet.D.J and Jepson.A. Stability of Phase information. *IEEE workshop on Visual Motion, Princeton.*, 1991.
- [8] Fleet.D.J and Jepson.A.D. The computation of normal velocity from local phase information. *IJCV*, 5:77–104, 1990.
- [9] Franks.L. *Signal Analysis*. Prentice-Hall,Inc,N.J., 1969.
- [10] Girod.B and Kuo.D. Direct Estimation of Displacement Histograms. *Proc. Image Understanding and Machine Vision.*, pages 73–76, 1989.
- [11] Heeger.D.J. A model for the extraction of image flow. *J.Opt.Soc.Am*, 4:1455–1471, 1987.
- [12] Horn.B.K.P and Schunk.B.G. Determining optic flow. *Artificial Intelligence*, 17:185–204, 1981.
- [13] Jasinschi.R and Rosenfield.A. Sumi.K. The perception of visual motion coherence and transparency: a statististical model. Technical report, Univ. Maryland, 1990.
- [14] Jepson.A.D. and Fleet.D.J. *Scale-Space Singularities*. (Ed.) Faugeras.O, Proc. ECCV,Antibes , Springer-Verlag, 1990.

- [15] Langley.K., Atherton.T.J., Wilson.R.G., and M.H.E.Larcombe. *Vertical and Horizontal Disparities from Phase*. (Ed.) Faugeras.O, Proc. ECCV,Antibes , Springer-Verlag, 1990.
- [16] Langley.K and Fleet.D.J. Causal velocity mechanisms applied to transparent surfaces. *Perception*, 20:79, 1991.
- [17] Langley.K and Fleet.D.J. Using group phase to solve the aperture problem. *Submitted to ECCV91*, 1991.
- [18] Langley.K, Rogers.B.J, and Brady.J.M. The Computation of Deformation and Rotation in Stereopsis. *BMVC90, University of Oxford*, pages 341–346, 1990.
- [19] Marr.D. *Vision*. W.H.Freeman & Co., 1982.
- [20] Papoulis.A. *Systems and Transforms with Applications in Optics*. McGraw-Hill, New York, 1968.
- [21] Shizawa.M and Mase.K. Determining multiple optic flow using spatio-temporal filters. *ICPR90, atlantic City*, pages 274–278, 1990.
- [22] Shizawa.M and Mase.K. Principle of Superposition: A common Computational Framework for Analysis of Multiple Motion. *IEEE workshop on visual motion,Princeton,N.J.*, 1991.
- [23] Shizawa.M and Mase.K. A unified computational theory for motion transparency and motion boundaries based on eigenenergy analysis. *CVPR91, Hawaii*, pages 289–294, 1991.
- [24] Uras.S, Girosi.F, Verri.A, and Torre.V. A Computational Approach to Motion Perception. *Biological Cybernetics*, 60:79–87, 1988.
- [25] Wang.J. A theory of image matching. *IEEE ICCV*, pages 200–209, 1990.

A Tensor Product Formulation of Constraints

To show that our approach can be viewed as formerly equivalent to that of Shizawa and Mase[21, 23], we form two constraint equations based upon instantaneous frequency:

$$\Omega \mathbf{V}_1^T \mathbf{V}_2 \Omega^T = 0 \quad (21)$$

$$\Omega \mathbf{V}_2^T \mathbf{V}_1 \Omega^T = 0 \quad (22)$$

where $\Omega = [k_{1i}, k_{2i}, \omega_i]^T$ is a $3 \times n$ matrix for $i = 1$ to n and $\mathbf{V}_1 = [u_1, u_2, 1]^T$, $\mathbf{V}_2 = [v_1, v_2, 1]^T$ and n is the number of independent measurements. The addition of the two constraints gives:

$$\Omega \mathbf{V}_m \Omega^T = 0 \quad (23)$$

with:

$$\mathbf{V}_m = \frac{1}{2}[\mathbf{V}_1 \otimes \mathbf{V}_2 + \mathbf{V}_2 \otimes \mathbf{V}_1] \quad (24)$$

using the identity:

$$\text{vec}(\Omega^T \mathbf{V}_m \Omega) = (\Omega^T \otimes \Omega^T) \text{vec}(\mathbf{V}_m) \quad (25)$$

where \otimes denotes the Kronecker product, and $\text{vec}(\mathbf{V}_m)$ refers to the 9×1 vector obtained by stacking the elements of \mathbf{V}_m in sequence, we obtain a least squares estimate of $\text{vec}(\mathbf{V}_m)$ using:

$$(\Omega \otimes \Omega)(\Omega^T \otimes \Omega^T) \text{vec}(\mathbf{V}_m) = (\Omega \Omega^T \otimes \Omega \Omega^T) \text{vec}(\mathbf{V}_m) = 0 \quad (26)$$

The matrix $(\Omega \otimes \Omega)(\Omega^T \otimes \Omega^T)$ has dimensions 9×9 , and is symmetric and with rank 5, in the case of two unambiguous flow fields while the column vector:

$$\text{vec}(\mathbf{V}_m) = [u_1 u_2, \frac{(u_1 v_2 + u_2 v_1)}{2}, \frac{u_1 + u_2}{2}, \frac{v_1 u_2 + v_2 u_1}{2}, v_1 v_2, \frac{v_1 + v_2}{2}, \frac{u_1 + u_2}{2}, \frac{v_1 + v_2}{2}, 1]^T \quad (27)$$

can be solved for using the properties of symmetry to reduce the eigensystem from 9 to 6 in a similar manner to Shizawa and Mase. Thus our approach is directly equivalent to that of Shizawa and Mase, but derived entirely from first order filter kernels.

B An Alternative Least-Squares Minimization

As an alternative to the eigenvalue problem discussed above with respect to (13) through (15), here we show a similar least-squares problem solved using the pseudo-inverse of the normal equations. We rewrite (13) through (15) by breaking \mathbf{m} into the form (\mathbf{A}, ω^2) where ω^2 is the last element of \mathbf{m} . Then (13) becomes

$$\mathbf{x}^T \mathbf{A} + \omega^2 = 0, \quad (28)$$

where $\mathbf{A} = (k_1^2, k_2^2, k_1 k_2, k_1 \omega, k_2 \omega)^T$, $\mathbf{x} = (u_1 u_2, v_1 v_2, u_1 v_2 + u_2 v_1, u_1 + u_2, v_1 + v_2)^T$. In this case, the least-squares solution to the five elements of \mathbf{x} minimizes the squared error:

$$\sum_j (\mathbf{x}^T A_j + \omega_j^2)^2 . \quad (29)$$

The corresponding normal equations are then solved by:

$$\mathbf{x} = (\tilde{M}^T \tilde{M})^{-1} \tilde{M}^T \mathbf{w} , \quad (30)$$

where the columns of \tilde{M}^T are A_j , and the j th element of \mathbf{w} is ω_j^2 .

In order to include the energy weighting on the instantaneous frequencies we include the diagonal weight matrix, $E = \text{diag}[E_1, \dots, E_n]$, where the diagonal entries represent the energy response from each quadrature filter pair used to estimate instantaneous frequency. Then the normal equations become:

$$\mathbf{x} = (\tilde{M}^T E \tilde{M})^{-1} \tilde{M}^T E \mathbf{w} , \quad (31)$$